减轻 CMT 中乱序程度的发送端数据分配方案 *

王振朝 [1,2]，李海潇 [1]，侯欢欢 [1]

(1. 河北大学 电子信息工程学院，河北 保定 071002; 2. 北京交通大学海滨学院，河北 沧州 061100)

摘　要：为减轻多径并行传输(concurrent multipath transfer，CMT)系统中的接收端数据包乱序对系统传输性能的影响，提出一种新的发送端数据分配方案。方案基于路径带宽、往返传输时延和拥塞窗口预测数据包前向传输时延，并将其作为划分系统中路径传输优先级的度量因子，发送端根据路径传输优先级和发送缓存区状态为各路径分配待发送队列中不会导致接收端乱序的数据包。仿真结果表明，与轮询和 ATLB 算法相比，文中所提发送端数据分配方案可有效减少接收端乱序数据包个数。

关键词：多径并行传输；数据乱序；数据分配；前向传输时延；吞吐量

中图分类号：TP393　　doi: 10.19734/j.issn.1001-3695.2018.07.0551

Data allocation algorithm for reducing disorder degree in cmt system

Wang Zhenchao[1, 2], Li Haixiao[1], Hou Huanhuan[1]

(1. *College of Electronic&Information Engineering*，*Hebei University*, *Baoding Hebei* 071002, *China*; 2. *Beijing Jiaotong University Haibin College*, *Changzhou Hebei* 061100, *China*)

Abstract: In order to reduce the effect of the packet disorder on the transmission performance at the receiver in the concurrent multipath transfer(CMT) system, this paper developed a new data allocation scheme in the transmission terminal. The scheme predicted the packet forward transmission delay based on the bandwidth, round trip transmission delay and congestion window of the path, and used it as a measure factor of the path transmission priority in the CMT system, the sender assigned packets in the queue for each path based on the path transmission priority and sender cache status, which do not cause the receiver to be out of order. Simulation results show that compared with the Round-Robin and ATLB algorithm, the data allocation scheme in this paper can reduce the number of disorder data apparently.

Key words: cmt; data disorder; data allocation; forward transmission delay; throughput

0　引言

无线接入技术 (radio access technologies,RAT)的不断发展以及终端可配备接口的不断增多，使得终端同时通过多种接入方式并行传输数据成为可能[1,2]。基于多接口技术的要求，IETF 工作组在 2009 年提出的多径并行传输协议 MPTCP（multipath transfer control protocol）[3,4]可在兼容 TCP 协议的基础上，为终端用户提供并行数据传输服务，有效地提高了 CMT 系统的传输能力和稳定性。但是多条路径在时延、带宽、丢包率等方面存在的差异，使得经多条路径传输的数据包有可能乱序到达接收端，导致接收缓存阻塞、系统吞吐量降低。

现有减轻接收端数据乱序程度的发送端数据分配方案，在数据包分配粒度的选择上主要有两种：批量分配和单个数据包分配。Xue 等人[5]提出一种基于动态反馈的前向预测数据包调度算法 DPSAF（forward prediction based dynamic packet scheduling and adjusting with feedback），考虑路径丢包率、发送时间偏差值对发送端数据分配的影响，计算在接收端不发生乱序的前提下应为每条路径分配的数据包个数，采用批量数据分配粒度将 DSN 编号连续的数据包分配给各路径。佘东平等人提出一种发送端数据分配方案[6]，通过实时估计各条路径吞吐量来评估路径的发送能力，发送端采用单个数据包分配粒度将数据包分配给发送能力最强的路径，从而达到减轻接收端乱序程度的目的。文献[5]采用批量分配粒度为各路径分配数据包，虽能有效提高发送端数据分配速率，但在路径质量相差较大的环境下，采用批量分配粒度将会加重接收端的乱序程度；文献[6]使用的单个数据包分配粒度，虽在一定程度上降低了发送端数据分配速率，但在分配的精确度上高于批量分配。由此可知，两种数据包分配粒度在分配速率和分配精度上的不同，使得其在不同的 CMT 系统环境中，对接收端乱序程度的影响不同，因此本文通过分析上述两种分配粒度的优劣，选择一种发送端数据分配粒度。

在选择出发送端数据分配粒度之后，如何将数据包分配到 CMT 系统中的各个路径上传输才能减轻接收端乱序程度是本文发送端数据分配方案的又一研究重点。针对数据包经不同路径传输乱序到达接收端的问题，已有许多学者提出了多种解决方案。Hasegawa 等人[7]提出一种基于分组到达时间的负载均衡算法 ATLB(arrival-time matching load-balancing)，预测数据包在每条路径上的前向传输时延，并将预测所得时延值定义为各路径的分数，分配数据包给得分最小的路径。但是算法在预测前向传输时延时，忽略了数据包发送时延的影响，使得时延预测值不准确。Chen 等人[8]提出一种发送端数据分配方案，考虑数据包发送时延和接收缓存区大小的限制为各路径分配数据包，使得接收端乱序问题得到一定程度

收稿日期：2018-07-24；**修回日期：**2018-09-07　**基金项目：**河北省自然科学基金资助项目（F2014201168）

作者简介：王振朝（1958-），男，山东鄄城人，教授，博士，主要研究方向为移动通信和物联网（wangzhenchaohbdx@163.com）；李海潇（1994-），女，河北廊坊人，硕士研究生，主要研究方向为并行多径传输中数据分配算法；侯欢欢（1990-），女，河北保定人，硕士，主要研究方向为并行多路径传输.

缓解。但上述两种算法在数据分配时均没有考虑路径拥塞窗口大小的限制，使得在各路径时延相差较大的环境下，会出现时延较小路径负载过重而其他路径空闲的情况。杜文峰等人[9]考虑路径时延、带宽、接收缓存区大小对发送端数据分配的影响，提出一种基于乱序反馈的差异化多径并行传输数据分配算法，建立路径性能评估模型得到路径性能参数，根据所得参数动态调节各路径传输数据包次数。黄辉[10]提出一种基于权重的数据调度算法，根据时延和丢包率定义路径质量，发送端选择质量最好的路径传输。文献[9，10]均是通过增加使用性能较好路径、减少或抑制性能较差路径传输数据包的次数来降低传输性能差的路径对系统整体传输性能的影响，但是此种方法降低了系统的资源利用率，浪费系统有效带宽。

为充分利用 CMT 系统有效带宽，实现负载均衡并降低接收端乱序程度，本文在上述算法的基础上提出了一种新的发送端数据分配方案。方案考虑路径前向传输时延和可发送窗口空闲值对发送端数据分配的影响，首先在文献[7，8]数据包前向传输时延预测方法的基础上考虑拥塞窗口的影响，预测数据包在各路径上的前向传输时延值，然后将预测所得时延值作为度量各路径传输优先级的因子；根据路径传输优先级和可发送窗口空闲值 $(CWND_i - Q_i)$ 的大小为 CMT 系统中各路径计算应分配的不会导致接收端乱序的数据包 DSN 编号，使得数据包按序到达接收端的同时，系统中有效带宽得到充分利用，从而提高系统资源利用率。

1　CMT 系统传输模型

多径并行传输数据分流可以在多个协议层（应用层、传输层、网络层以及链路层）实现[11]，本文研究传输层的多径并行传输数据分流，建立如图 1 所示的 CMT 系统传输模型。应用层中的数据流到达传输层之后，传输层按照一定的负载分割粒度将数据流分割成大小为 Datasize 的数据包，并在每个数据包首部添加数据层序列编号(data sequence number，DSN)（用于接收端的数据重组）和子流层序列编号(subflow sequence number，SSN)（避免不必要的快速重传）。

在现有 CMT 系统中，各条子路径共享一个发送缓存区，这样系统很难充分利用质量较好路径的传输能力，而且还会因为缓存区空间有限而造成传输阻塞[12]，因此，本文为各条 TCP 子路径分配独立的发送缓存区。若路径发送缓存区中可发送窗口存在空闲，则为该路径分配的数据包能直接传输到路径上；若没有空闲，为该路径分配的数据包则需等待其中的数据包成功到达接收端，窗口重新出现空闲才能传输到路径上。可发送窗口受拥塞窗口状态变量控制，大小等于路径拥塞窗口值，并随拥塞控制策略动态变化，因此为各路径分配独立的发送缓存区可使得质量较好路径承担更多的数据传输任务，从而提高 CMT 系统的数据传输速率。

在上述模型的基础上本文主要解决如下问题：发送端根据一定的条件判断出要将 DSN 编号为 k 的数据包（数据包 k）分配给路径 i 传输才能使得系统中已发送数据包按序到达接收端，而路径 i 可发送窗口此时填满已发送但未被确认的数据包，其他路径可发送窗口存在空闲。现有发送端数据分配算法[13~18]在此种情况下主要有两种做法：a）为避免接收端数据包乱序到达，尽管其他路径存在可用带宽也不利用，而是等待该路径可发送窗口重新出现空闲再将 DSN 编号为 k 的数据包发送出去。该方法虽能有效减少接收端的乱序程度，但是浪费系统有效带宽，降低发送端数据传输速率；b）直接将 DSN 编号为 k 的数据包分配给其他路径，该方法虽能有效利用系统中可用带宽，但在时延相差较大环境下有可能导致接收端严重乱序。

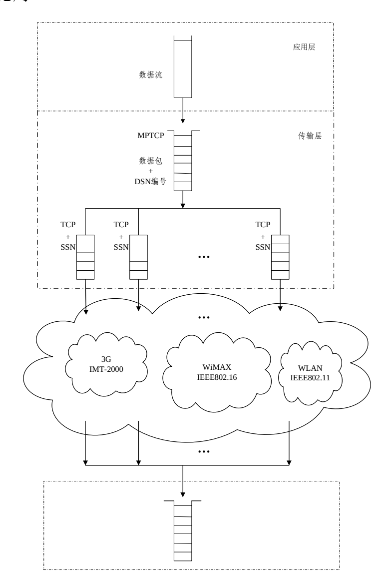


图 1　CMT 系统传输模型

Fig. 1　Transmission model of CMT system

本文提出的发送端数据分配方案在上述背景的基础上，改变原有发送端数据包分配模式，如图 2 所示，不必等待数据包 k 发送出去之后才将发送端中 DSN 编号大于 k 的数据包发送到路径上，而是根据路径传输优先级和可发送窗口剩余量计算 N_x，取发送端中数据包 $(k + N_x)$ 分配给可发送窗口空闲的路径 x，从而在避免接收端乱序的情况下充分利用系统有效带宽。

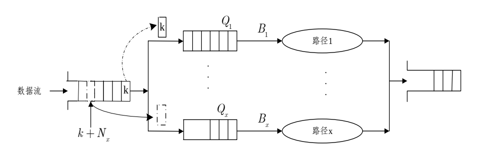


图 2　数据分配方式

Fig. 2　Data allocation scheme

2　发送端数据分配方案

2.1　路径传输优先级的划分

本文发送端数据分配方案通过预测数据包在各条路径上的前向传输时延值为各路径划分传输优先级。数据包在路径 i 上的前向传输时延值 T_i 可用下式求得：

$$T_i = \begin{cases} \dfrac{Q_i}{B_i} + \dfrac{Datasize}{B_i} + \dfrac{RTT_i}{2}, Q_i < CWND_i \\ \dfrac{CWND_i}{B_i} + \dfrac{Datasize}{B_i} + \dfrac{RTT_i}{2} + \dfrac{RTT_i}{2}, Q_i \ge CWND_i \end{cases} \quad (1)$$

其中：Q_i 表示路径 i 发送缓存区中已有的数据长度，B_i 表示路径 i 的带宽，Datasize 表示发送端数据包大小，RTT_i 表示路径 i 的往返传输时延，$CWND_i$ 表示路径 i 的拥塞窗口值。

当 $Q_i < CWND_i$ 时，表示路径 i 可发送窗口存在空闲，数据包分配到该路径之后，只需等待缓存区中已有数据发送到路径上即可进行发送。式中第一项表示发送缓存区中已有数据发送出缓存区所需的时间，第二项表示预测的数据包发送出缓存区所需的时间，第三项表示预测的数据包发送出缓存区之后到到达接收端所需的时间；当 $Q_i \ge CWND_i$ 时，表示路径 i 发送缓存区可发送窗口中填满已发送但未被确认的数据包，数据包分配到路径 i 发送缓存区之后，需等待可发送窗口中

的已有数据包成功确认到达接收端并释放出缓存区，才能发送到路径中。因此式中第一、二、三项表达式意义同上，第四项表示接收端返回确认接收消息 ACK 所需的时间。

根据预测所得时延值为各路径划分传输优先级，数据包在某条路径上的前向传输时延值越大，表示数据包在该路径上传输到接收端所需的时间越长，则路径传输优先级越低。因此将路径 i 的传输优先级定义为

$$PP_i = \frac{1}{T_i} \tag{2}$$

2.2　发送端数据包分配粒度的确定

划分路径传输优先级之后，发送端需确定数据包的分配粒度。现有发送端数据分配算法的数据分配粒度主要有两种[5,6，11~15]：批量分配和单个数据包分配。

批量分配粒度指发送端取 DSN 编号连续的数据包填满选定路径发送缓存区，这种分配粒度虽大大增加了发送端的数据分配效率，但是对系统路径传输性能要求严格，在路径性能相差较大的环境下，会导致严重的接收端乱序现象。如图 3 中发送端采用批量分配粒度为各路径分配数据包，若要保证接收端不发生数据包乱序到达现象，则必须保证路径 j 上第一个数据包晚于路径 i 上第 $CWND_i$ 个数据包到达接收端，才能避免接收端数据包乱序，即需满足 $CWND_i \frac{Datasize}{B_i} + \frac{RTT_i}{2} < \frac{Datasize}{B_j} + \frac{RTT_j}{2}$，由此可知，批量分配粒度对系统路径性能要求严格。

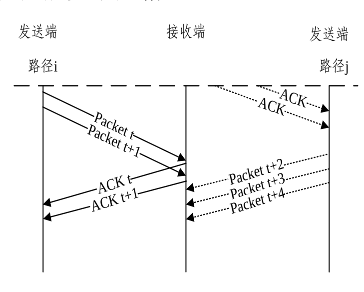


图 3　批量分配粒度

Fig. 3　Batch allocation

而单个数据包分配方式虽相比批量分配来说降低了数据分配效率，但却提高了数据分配精确度。如图 4 中发送端采用单个数据包分配粒度为路径分配数据包，在为路径 i 分配一个数据包之后便更新路径参数，重新衡量路径传输性能，继而选择下一次数据分配路径 j,依此类推。单个数据包分配粒度对路径性能要求不像批量分配那样严格，适用于各种路径性能下的 CMT 系统，所以本文采用单个数据包分配粒度为各路径分配数据。

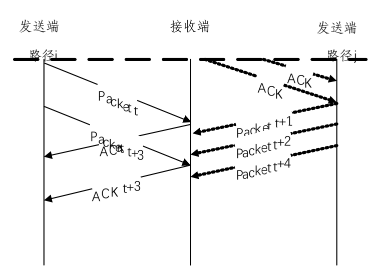


图 3　单个数据包分配粒度

Fig. 3　Single data allocation

2.3　各路径传输数据包 DSN 编号的确定

确定路径传输优先级和发送端数据分配粒度之后，本文发送端数据分配方案如图 2 所示，CMT 系统中路径传输优先级从高到低排序得 $\{P_1, P_2, \ldots, P_n\}$。

若 $Q_1 < CWND_1$，则表示传输优先级最高路径可发送窗口存在空闲，将此时发送端中 DSN 编号最小的数据包 k 分配给该路径，然后更新路径 1 的传输优先级因子，重新选择优先级最高路径。

若 $Q_1 \geq CWND_1$，传输优先级最高路径可发送窗口空闲值为 0，此时发送端分配给该路径的数据包不能发送到路径上，但其他可发送窗口空闲的路径具有数据包发送能力。按传输优先级从高到低顺序找出路径 P_x，满足 $Q_x < CWND_x, 1 \leq x \leq n$，发送端根据式（3）~(6）为路径 x 计算应分配的数据包 DSN 编号。

路径 x 的传输优先级为 PP_x，将此时发送端中最小 DSN 编号 k 之后的第 N_x 个数据包分配给该路径，N_x 计算过程如下：

$$N_x = N_1 + N_2 \tag{3}$$

其中:

$$N_1 = \sum_{i=1}^{k}(CWND_i - outsding_i) + \frac{T_x - (T_i + FT_i)}{d_{trans_i}} \tag{4}$$

$$N_2 = \sum_{i=k+1}^{x-1} \frac{T_x - T_i}{d_{trans_i}} \tag{5}$$

其中: $k = \arg\max_i^x (T_x - T_i < FT_i)$。分配完成之后，更新路径的传输优先因子，重新排列传输优先级，重复上述步骤选择路径分配数据包，直至发送端没有可发送数据包为止。

2.4　数据分配流程图

本文发送端数据分配方案通过为系统中可发送窗口空闲的路径分配合适 DSN 编号的数据包，在保证数据包按序到达接收端的前提下，充分利用系统有效带宽，从而提高系统吞吐量和数据传输速率。数据分配流程如图 5 所示。

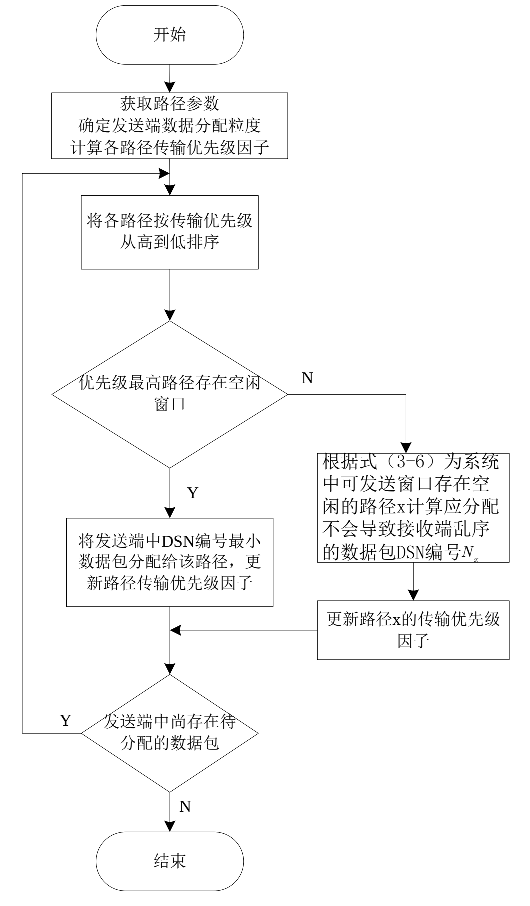


图 5　数据分配流程图

Fig. 5　Data allocation flow chart

3 仿真结果及分析

3.1 仿真环境及参数设置

本文采用 Linux 操作系统下的 NS-3 网络仿真器[20]来验证本文所提算法的有效性。仿真实验的通信协议采用支持多径并行传输的 MPTCP 协议。数据包的大小为 1500 Byte，重传策略选择 RTX_LOSSRATE,接收端缓冲区大小为 2.55 MB[19]。应用层协议使用 FTP 协议传输一个足够大的文件（即在整个仿真过程中 FTP 数据流一直持续）来模拟实际传输的过程。

构建如图 6 所示的网络拓扑环境，CMT 系统的发送端和接收端之间存在 3 条异构网络并行传输路径。设置如下所示场景对本文所提发送端数据分配方案和轮询、ATLB 算法进行仿真，路径 1、2、3 的带宽分别为 1 Mbps、1.5 Mbps 和 2 Mbps；往返传输时延分别为 50 ms、50 ms 和 500 ms。

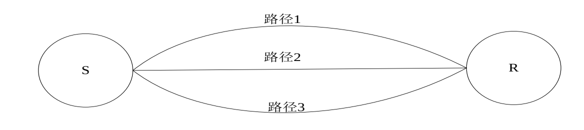


图 6 仿真网络拓扑图

Fig. 6 Simulation network topology

3.2 仿真结果分析

图 7 给出了发送端使用轮询、ATLB 算法和本文方案在上述场景下为各路径分配数据包的接收端数据包乱序情况。由图 7 可知，发送端使用三种分配方案为各路径分配数据包，接收端均存在不同程度的乱序。使用轮询算法并行传输数据过程中，接收缓存区中大部分时间存在 16~18 个乱序的数据包，使用 ATLB 算法接收端缓存中大部分时间存在 9~11 个乱序数据包，而使用本文方案接收缓存中大部分时间存在 4~6 个乱序数据包。由此可知本文方案相较于轮询和 ATLB 算法而言，乱序到达接收端的数据包个数明显减少。其主要原因是，轮询算法受性能较差路径影响较大，使得数据包大量乱序到达接收端；ATLB 算法性能虽优于轮询算法，但其通过预测数据包前向传输时延为各路径分配数据包，时延预测值的不准确导致数据包乱序到达接收端；而本文方案可以很好的根据路径性能计算出应分配给各路径的不会导致接收端乱序的数据包 DSN 编号，使得数据包能顺序到达接收端，并顺利递交给应用层。

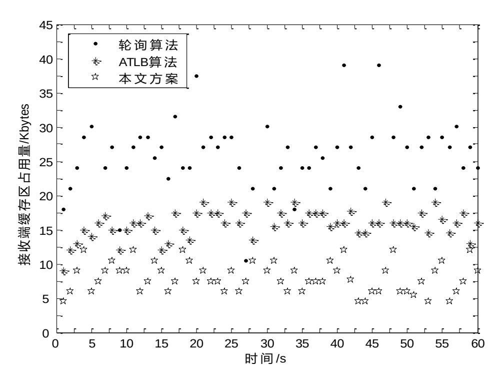


图 7 接收端缓存占用情况

Fig. 7 Receiver cache occupancy

改变路径 3 的往返传输时延值，验证轮询、ATLB 算法和本文方案在不同路径时延差场景下的的接收端缓存占用情况，如图 8 所示，发送端使用本文方案分配数据包接收端平均乱序数据包个数均小于轮询和 ATLB 算法。其主要原因是轮询算法不考虑路径质量依次为各路径分配数据包，路径时延差的增加使得接收端乱序数据包个数随之增加；ATLB 算法根据预测所得时延值为时延最小的路径分配数据包，随着路径时延差的增加，CMT 系统中将会出现时延较小路径负载过重而时延较大路径空闲的情况，此时，ATLB 算法在预测数据包在时延较小路径上的前向传输时延时，既忽略了数据包发送时延的影响又忽略了接收端 ACK 信息返回时间的影响，导致数据包分配不准确，接收端乱序数据包个数增加。因此可知，本文方案在路径性能相差较大和较小环境下，均能很好的为各路径分配数据包，有效减少接收端乱序数据包个数，提高系统吞吐量。

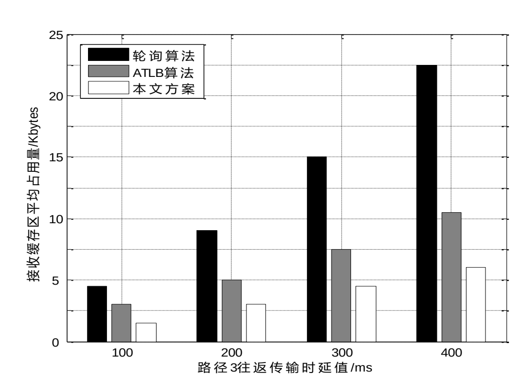


图 8 接收端缓存占用情况对比

Fig. 8 Comparison of the cache occupancy

通过仿真分析可知，发送端使用本文方案为各路径分配数据，可有效减少接收端乱序数据包个数，从而提高系统吞吐量。

4 结束语

本文首先通过分析数据包在 CMT 系统中传输直至递交到上层应用所经历的过程，建立一个传输模型，并在此模型基础上预测数据包在各条路径上的前向传输时延值，并将所得预测值作为划分各路径传输优先级的度量因子；然后发送端根据传输优先级和发送缓存区中可发送窗口空闲值的状态为各路径按单个数据包分配粒度分配合适 DSN 编号的数据包，以保证数据按序高效到达接收端。通过 NS-3 仿真验证得出，使用本文发送端数据分配方案为各路径分配数据包，可有效减少接收端乱序数据包个数，从而提高系统的数据传输速率和吞吐量。

参考文献:

[1] Verma L P, Kumar M. An adaptive data chunk scheduling for concurrent multipath transfer [J]. Computer Standards&Interfaces, 2017, 52(33): 97-104.

[2] Huang C J, Guan C T, Chen Hengming, *et al.* A self-adaptive joint bandwidth allocation scheme for heterogeneous wireless networks [J]. Applied Soft Computing, 2015, 37 (2015): 156-165.

[3] Oh BH, Lee JY. Constraint-based proactive scheduling for MPTCP in wireless networks [J]. Computer Networks, 2015, 91 (2015): 548-563.

[4] Li Xiaokai, Zhang Xiangli. Aconcurrent multi-path transmission of data allocation algorithm based on heterogeneous networks[C]//Proc of the 3rd International Conference on Systems and Informatics.2016: 487-491.

[5] Xue Kaiping, Han Jiangping, Ni Dan, *et al.* DPSAF: forward prediction based dynamic packet scheduling and adjusting with feedback for multipath TCP in lossy hererogeneous networks [J]. IEEE Trans on Vehicular Technology, 2018, 67(2): 1521-1534.

[6] 余东平，王聪，张剑锋,等．多路并行传输中数据调度机制的研究[J]．军事通信技术, 2015, 36(3): 18-22．（Yu Dongping, Wang Cong, Zhang Jianfeng, *et al.* Research on data scheduling mechanism in concurrent multipath transfer [J]. Journal of Military Communications Technology, 2015, 36 (3): 18-22．）

[7] Hasegawa Y, Yamaguchi, Hama T, *et al.* Improved data distribution for multipath TCP communication[C]//Proc of IEEE Global Telecommunications Conference. 2005: 271-275.

[8] Chen Rui, Wang Wennai, Zhu Jingling, *et al.* A round-trip-time based concurrent transmission scheduling for MPTCP[C]//Proc of the 6th International Conference on Wireless Communications and Signal Processing. 2014: 179-183.

[9] 杜文峰，吴真．基于乱序反馈的差异化多路径并发传输模型数据分配算法 [J]．计算机科学, 2015, 42(3): 60-64.（Du Wenfeng, Wu Zhen. Data distribution algorithm with out-of-order feedback for CMT over diversity network [J]. Computer Science, 2015, 42(3): 60-64．）

[10] 黄辉．基于权重的MPTCP数据调度算法设计 [J]．软件, 2016, 37(2): 77-80．（Huang Hui. Weight-based Data Scheduling Algorithm Designing for MPTCP [J]. Computer Engineerin&Software, 2016, 37(2): 77-80．）

[11] Nakasan C, Ichikawa K, Uthayopas P. Performance evaluation of MPTCP over open flow network [J]. Information Processing Society of Japan SIG Notes, 2014, 2014 (30): 1-6.

[12] Zhou Dizhi, Song Wei, Wang Ping, *et al.* Multipath TCP for user cooperation in LTE networks [J]. IEEE Network Magazine, 2015, 29 (1): 18-24.

[13] 王振朝，侯欢欢，李海潇．并行多径传输中基于服务质量的路径优化方案 [J]．小型微型计算机系统，2018，39(3)：503-507．（Wang Zhenchao, Hou Huanhuan, Li Haixiao. Path optimization scheme based on quality of service in concurrent multipath transmission [J]. Journal of Chinese Computer Systems, 2018, 39(3): 503-507．）

[14] Xu Changqiao, Liu Tianjiao, Guan Jianfeng, *et al.* CMT-QA: quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks [J]. IEEE Trans on Monile Computing, 2013, 12(11): 2193-2205.

[15] Prabhavat S, Ansari N, Kato N, *et al.* Effective delay-controlled load distribution over multipath networks [J]. IEEE Trans on Parallel and Distributed Systems, 2011, 22(10): 1730-1741.

[16] Oh B H, Lee J Y. Constraint-based proactive scheduling for MPTCP in wireless networks [J]. Computer Network, 2015, 91(14): 549-563.

[17] Gong Weibing, Yang xiaolong, Zhang Min, *et al.* An adaptive traffic distribution scheme for CMT based on Lotka-Volterra model in multihomed networks [J]. China Communications, 2017, 14(2): 79-89.

[18] Xu Changqiao, Li Zhuofeng, Zhong Lujie, *et al.* CMT-NC: improving the concurrent multipath transfer using network coding in wireless networks [J]. IEEE Trans on Vehicular Technology, 2016, 65(3): 1735-1751.

[19] Chen Sengyang, Yuan Zhenhui, Muntean G M. An energy-aware multipath TCP-based content delivery scheme in hererogeneous wireless network[C]//Proc of IEEE Wireless Communication and Networking Conference. 2013: 1291-1296.

[20] MPTCP implementation for NS-3[EB/OL]. http://code. google. com/p/mptcp-ns. 3.